

Tilburg University

Signal Extraction and the Formulation of Unobserved Components Models

Harvey, A.C.; Koopman, S.J.M.

Publication date:
1999

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Harvey, A. C., & Koopman, S. J. M. (1999). *Signal Extraction and the Formulation of Unobserved Components Models*. (CentER Discussion Paper; Vol. 1999-44). Econometrics.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Signal Extraction and the Formulation of Unobserved Components Models

Andrew Harvey and Siem Jan Koopman
Faculty of Economics and Politics, University of Cambridge
and Center for Economic Research, Tilburg University

May 25, 1999

Abstract

This paper looks at unobserved components models and examines the implied weighting patterns for signal extraction. There are three main themes. The first is the implications of correlated disturbances driving the components, especially those cases in which the correlation is perfect. The second is how setting up models with *t-distributed* disturbances leads to weighting patterns which are robust to outliers and breaks. The third is a comparison of implied weighting patterns with kernels used in nonparametric trend estimation and equivalent kernels used in spline smoothing. We also examine how weighting patterns are affected by heteroscedasticity and irregular spacing and provide an illustrative example.

KEYWORDS: Cubic spline, Kalman filter and smoother, Kernels, Robustness, Structural time series model, Trend, Wiener-Kolmogorov filter.

JEL classification: C15, C22.

1 Introduction

Once the parameters in an unobserved components (UC) model have been estimated, interest often centres on estimates of the components themselves. These components usually have a direct interpretation. For example, a component may represent an underlying trend, seasonal or cycle in the series. Identifiability of the components normally requires that some assumption be made about the correlation between the disturbances driving them. The most common assumption is that they are mutually uncorrelated, though it is sometimes argued that models with perfectly correlated disturbances have certain attractions; see Snyder (1985) and Ord, Koehler and Snyder (1997). Models with perfectly correlated disturbances also arise as a consequence of the decomposition of Beveridge and Nelson (1981).

Assumptions made about the correlations between disturbances are rarely suggested by prior knowledge of the components with which they are associated. Instead the choice is governed by statistical considerations. One of these considerations is the implicit way in which the observations are weighted in order to extract a component. The main purpose of section 2 is to examine the implications of different assumptions about the correlation between the disturbances in a random walk plus noise model. The results are established analytically using the Wiener-Kolmogorov (WK) filter. Some interesting side issues arise as a consequence of this investigation. These include the contrast between two forms of the state space model for correlated disturbances and the weighting pattern for a model in continuous time. Section 3 derives the weighting pattern for the local linear trend model with uncorrelated disturbances. This is used later in connection with cubic splines.

Section 4 looks at two methods which have been suggested in the literature for extracting a trend from an autoregressive integrated moving average (ARIMA) model. One is the canonical decomposition of Hillmer and Tiao (1982). The other is the Beveridge-Nelson (BN) decomposition. The trend in the BN decomposition is obtained from a one-sided filter, but it is possible to construct a corresponding two-sided filter.

Section 5 discusses the consequences of heteroscedasticity and irregular spacing for signal extraction. In these cases the weights cannot be obtained from the WK formula, but they can be computed numerically using the Kalman filter and smoother (KFS) algorithm given in appendix A.

The filters considered in section 2 are all linear and are optimal, in the sense of minimising mean square estimation error, within the class of linear estimators. However, unless the disturbances are Gaussian, they are not optimal within the class of all possible estimators. Durbin and Koopman (1997) show how to estimate the parameters in non-Gaussian UC models using simulation techniques and how to estimate the components. Section 6 examines the nonlinear signal extraction filters which emerge for such models and illustrates the form they take when an outlier or structural break is present. Heavy-tailed disturbance distributions yield robust estimates in such cases and the emphasis is on models in which the distributions are Student's t .

Section 7 makes some observations regarding the relationship between nonparametric estimates of the trend and the implied weighting patterns for UC models. The implied weighting patterns can be viewed as kernels and it is interesting to compare the shape of model-based kernels with those typically used in nonparametric work and to explore the relationship of signal-noise ratios to bandwidth. It is also shown how the implied weighting patterns for splines can be obtained by relating them to UC models.

In section 8 we illustrate many of the points made about signal extraction by fitting a UC time series model to the motorcycle acceleration data given in Silverman (1985). Silverman took a nonparametric approach to fitting a cubic spline. We fit a cubic spline based on a UC model and show that the weighting patterns implied by the two approaches differ when the observations are not equally spaced. We also use the local level model to analyse the data and make comparisons with the cubic spline using the Akaike information criterion (AIC). Finally, we propose a simple method of allowing for heteroscedasticity in the data.

Section 9 concludes by arguing that the weighting of observations is fundamental to prediction and signal extraction, and that UC models provide a framework for determining the best way in which weights should be constructed.

2 Correlated Components

The simple random walk plus noise, or local level, model can be used to explore the implications of correlated disturbances. The model can be written as follows:

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \quad t = 1, \dots, T, \quad (1)$$

$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim WN(0, \alpha^2 \sigma_\varepsilon^2), \quad (2)$$

where WN denotes ‘white noise’, that is serially uncorrelated random variables, and

$$E(\varepsilon_t \eta_s) = \rho \alpha \sigma_\varepsilon^2, \quad t = 1, \dots, T, \quad E(\varepsilon_t \eta_s) = 0, \quad t \neq s. \quad (3)$$

with $|\rho| \leq 1$ and $\alpha \geq 0$. An alternative model has the transition equation written at time t , that is

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim WN(0, \alpha^2 \sigma_\varepsilon^2), \quad (4)$$

with the covariance still as in (3). When the correlation, ρ , is zero the two models are essentially the same. For non-zero ρ the weighting patterns display an interesting asymmetry. In what follows we will refer to (2) as the *future* state form and (4) as the *contemporaneous* state form.

The parameters of the ARIMA(0,1,1) reduced form are found by first differencing. In the case of the future state model this yields

$$\Delta y_t = \eta_{t-1} + \varepsilon_t - \varepsilon_{t-1} = \xi_t + \theta \xi_{t-1}, \quad \xi_t \sim WN(0, \sigma^2), \quad t = 2, \dots, T, \quad (5)$$

and so the variance and first-order autocovariance are, respectively,

$$\gamma(0) = \sigma_\varepsilon^2(\alpha^2 + 2 - 2\rho\alpha) = \sigma^2(1 + \theta^2), \quad (6)$$

and

$$\gamma(1) = \sigma_\varepsilon^2(\rho\alpha - 1) = \sigma^2\theta. \quad (7)$$

The first-order autocorrelation is

$$\rho(1) = \frac{\rho\alpha - 1}{\alpha^2 + 2 - 2\rho\alpha} = \frac{\theta}{1 + \theta^2}, \quad (8)$$

from which

$$\theta = \frac{\alpha^2 + 2(1 - \rho\alpha) - \sqrt{\alpha^4 + 4\alpha^2(1 - \rho\alpha)}}{2(\rho\alpha - 1)}. \quad (9)$$

Choosing the negative sign in front of the square root gives the invertible reduced form parameter, that is $|\theta| < 1$, provided $\alpha > 0$ and we do not have $\rho = 1$ with $\alpha = 2$. If $\alpha = 0$, then $\theta = -1$, while $\rho = 1$ and $\alpha = 2$ gives $\theta = 1$. It is clear from (7) that, for positive σ_ε^2 , the sign of θ is given by the sign of $\rho\alpha - 1$. Thus for $\rho \leq 0$, θ is negative, while for $\rho > 0$, θ is only negative if $\rho\alpha < 1$.

The variance of the reduced form disturbance is obtained from (7) as

$$\sigma^2 = \sigma_\varepsilon^2(\rho\alpha - 1)/\theta, \quad \theta \neq 0. \quad (10)$$

Note that if $\rho\alpha = 1$, then $\theta = 0$ and $\sigma^2 = \alpha^2 \sigma_\varepsilon^2 = \sigma_\eta^2$. It is also the case that $\theta = 0$ when $\sigma_\varepsilon^2 = 0$, in which case $\alpha \rightarrow \infty$, and ρ is not defined.

When the transition equation is as in (4), the variance and first-order autocovariance are

$$\gamma(0) = \sigma_\varepsilon^2(\alpha^2 + 2 + 2\rho\alpha), \quad (11)$$

and

$$\gamma(1) = -\sigma_\varepsilon^2(\rho\alpha + 1). \quad (12)$$

The reduced form MA parameter, θ , is given by expression (9), but with ρ replaced by minus ρ . The sign of θ is now given by the opposite of the sign of $\rho\alpha + 1$. Thus for ρ positive, θ is negative, while for ρ negative, θ is only negative if $\alpha < -1/\rho$.

As (9) makes clear, there is an identifiability problem in that different values of ρ and α can give rise to the same θ . However, on substituting from (7) into (6) it can be seen that

$$\sigma_\eta^2 = \sigma^2(1 + \theta)^2. \quad (13)$$

The same result holds for the contemporaneous form. Thus σ_η^2 is always identified and so the identifiability problem can be seen in terms of σ_ε^2 and ρ rather than α and ρ . The usual way out of the problem is to set $\rho = 0$, but it could be set to any value. This section explores the implications of different values of ρ for filtering and signal extraction.

2.1 Filtering and prediction

The optimal linear predictor of a future observation in a model with an ARIMA(0,1,1) reduced form depends only on θ . Given observations up to and including y_t , the optimal linear predictor of y_{t+1} is

$$\tilde{y}_{t+1|t} = (1 + \theta) \sum_{j=0}^{\infty} (-\theta)^j y_{t-j} = \lambda \sum_{j=0}^{\infty} (1 - \lambda)^j y_{t-j}, \quad |\theta| < 1, \quad (14)$$

where $\lambda = 1 + \theta$ is the smoothing constant in the exponentially weighted moving average (EWMA). A negative θ corresponds to a value of λ between zero and one.

Now since the optimal linear predictor of ε_{t+1} , based on information at time t is zero, expression (14) also gives the one-step ahead filtered estimator of the level, that is $\tilde{\mu}_{t+1|t}$. With the level as in (4), this is also the filtered estimator of μ_t based on information at time t . However, unless ρ is zero, this is not the case when the level is formulated as in (2) since

$$\tilde{y}_{t+1|t} = \tilde{\mu}_{t+1|t} = \tilde{\mu}_{t|t} + \alpha \rho \tilde{\varepsilon}_t, \quad (15)$$

where $\tilde{\varepsilon}_t$ is the optimal linear predictor of ε_t based on information at time t .

Expression (14) gives the optimal forecast any number of steps ahead; it is the ‘eventual forecast function’. Having $\tilde{\mu}_{t|t}$ equal to the eventual forecast function is an attractive property. Its mean square error (MSE) is the price paid for not knowing the starting point of the eventual forecast function. Although, when the level is set up as in (4), $\tilde{\mu}_{t|t}$ is invariant to ρ , its MSE is not. We have

$$MSE(\tilde{y}_{t+1|t}) = MSE(\tilde{\mu}_{t|t}) + \sigma_{\eta}^2 + \sigma^2(2\rho\alpha + 1).$$

In the steady-state $MSE(\tilde{y}_{t+1|t}) = \sigma^2$, and so, substituting from (10),

$$MSE(\tilde{\mu}_{t|t}) = \sigma^2(1 + \theta)[- \theta - (1 + \theta)(\rho/\alpha)]. \quad (16)$$

2.2 Smoothing and signal extraction

The classical WK formula for finding the weights used to extract the minimum mean square linear estimator (MMSLE) of a component, μ_t , in a model of the form (1) is

$$\tilde{\mu}_t = \sum_{j=-\infty}^{\infty} w_{-j} L^j y_t = w(L) y_t, \quad w(L) = \frac{\gamma_{\mu}(L) + \gamma_{\mu\varepsilon}(L)}{\gamma_y(L)}, \quad (17)$$

where L is the lag operator, $\gamma_{\mu}(L)$ is the autocovariance generating function (ACGF) of μ_t and $\gamma_{\mu\varepsilon}(L)$ is the cross-covariance generating function of μ_t and ε_t . The ACGF of y_t is

$$\gamma_y(L) = \gamma_{\mu}(L) + \gamma_{\mu\varepsilon}(L) + \gamma_{\varepsilon\mu}(L) + \gamma_{\varepsilon\varepsilon}(L), \quad (18)$$

though this is usually evaluated in terms of the reduced form parameters. For a stationary autoregressive moving average (ARMA) process, written $\phi^{-1}(L)\theta(L)\xi_t$, where $\phi(L)$, and $\theta(L)$, are polynomials in the lag operator and ξ_t is white noise with variance σ^2 , the ACGF is given directly by

$$\gamma(L) = \{|\theta(L)|^2 / |\phi(L)|^2\} \sigma^2, \quad (19)$$

where $|\theta(L)|^2 = \theta(L)\theta(L^{-1})$, and similarly for $\phi(L)$.

Although formula (17) is only proved for stationary models in Whittle (1983, pp.56-58), it is argued in Bell (1984) and Burridge and Wallis (1988) that it can still be used for nonstationary models even

though expressions like (19) are no longer ACGFs. Thus for (1) and (2), we have, provided $\alpha > 0$ and we do not have $\rho = 1$ and $\alpha = 2$,

$$w(L) = \frac{\sigma_\varepsilon^2 \alpha^2 + \rho\alpha L(1 - L^{-1})}{\sigma^2 |1 + \theta L|^2} = \frac{\alpha\theta}{\rho\alpha - 1} \frac{\alpha - \rho + \rho L}{|1 + \theta L|^2} = \frac{(1 + \theta)^2}{\alpha} \frac{\alpha - \rho + \rho L}{|1 + \theta L|^2}, \quad (20)$$

the last equality following on noting that re-arranging (8) gives

$$\alpha^2\theta/(\rho\alpha - 1) = (1 + \theta)^2, \quad \rho\alpha \neq 1.$$

The formula in (20) is valid even if the noninvertible θ is chosen, since setting θ to $1/\theta$ gives the same result in view of the fact that $|1 + \theta^{-1}L|^2 = |1 + \theta L|^2/\theta^2$.

The weights in $w(L)$ sum to unity as can be seen immediately on setting $L = 1$ in (20). This is generally true for extracting any signal which contains a root equal to one, provided, of course, that there is no other component with such a root. The result follows from (17) once numerator and denominator, the latter expressed as (18), have been multiplied by $|1 - L|^2$.

On noting that, with $|\theta| < 1$, $|1 + \theta L|^{-2}$ is the ACGF of an AR(1) process with unit disturbance variance, (20) gives

$$\tilde{\mu}_t = \frac{1 + \theta}{1 - \theta} \left[\left(1 - \frac{\rho}{\alpha}\right) \sum_{-\infty}^{\infty} (-\theta)^{|j|} y_{t+j} + \frac{\rho}{\alpha} \sum_{-\infty}^{\infty} (-\theta)^{|j|} y_{t+j-1} \right], \quad (21)$$

or

$$\tilde{\mu}_t = \frac{1 + \theta}{1 - \theta} \left[\sum_{-\infty}^{\infty} (-\theta)^{|j|} y_{t+j} + \frac{\rho}{\alpha} \sum_{-\infty}^{\infty} (-\theta)^{|j|} \Delta y_{t+j} \right].$$

For $\theta \neq 0$, the weights can be expressed as

$$\begin{aligned} w_j &= \frac{1 + \theta}{1 - \theta} \left[1 - \frac{\rho}{\alpha} \left(1 + \frac{1}{\theta}\right) \right] (-\theta)^{-j}, & j \leq -1 \\ w_j &= \frac{1 + \theta}{1 - \theta} \left[1 - \frac{\rho}{\alpha} (1 + \theta) \right] (-\theta)^j, & j \geq 0 \end{aligned} \quad (22)$$

while for $\theta = 0$

$$w_{-1} = \rho^2, \quad w_0 = 1 - \rho^2, \quad w_j = 0, \quad j \neq -1, 0.$$

A value of $\theta = 0$ arises when $\rho\alpha = 1$ and when $\sigma_\varepsilon^2 = 0$ in which case $w_0 = 1$.

When $\rho = 0$ it can be seen from (21) that the weights in $w(L)$ decline symmetrically and exponentially, that is

$$w_j = \{(1 + \theta)/(1 - \theta)\}(-\theta)^{|j|}, \quad j = 0, 1, 2, \dots \quad (23)$$

When θ is negative, all weights will be positive if the terms in square brackets in (22) are positive. When ρ is positive, in which case negative θ requires $\rho\alpha < 1$, the term in square brackets is always positive (it is greater than one) for $j \leq -1$ because $\rho(1 + \theta)/\theta\alpha$ is negative. For $j \geq 0$, positive weights require $\rho(1 + \theta)/\alpha < 1$, but this can be shown to follow from (7) and (13). Since $w_{-j} > w_j$, $j = 1, 2, 3, \dots$, past observations receive relatively more weight. When ρ is negative future observations receive more weight. The weights for $j \leq -1$ are positive provided $\rho(1 + \theta)/\theta\alpha < 1$.

With θ positive, weights are negative and positive as $(-\theta)^{|j|}$ keeps changing sign. When $1 < \rho\alpha < 2$, past observations dominate in the sense that their absolute values are bigger. The weights are skew-symmetric when $\rho\alpha = 2$, that is $w_{-j} = -w_j$, $j = 1, 2, \dots$, and future observations dominate when $\rho\alpha > 2$.

Table 1 summarizes the above findings. Figure 1 gives some illustrations. The symmetry around a lag of one observed in the first panel, that is $w_{-j-2} = w_j$, $j = 0, 1, 2, \dots$, arises whenever $\alpha = \rho$. In the second panel, $w_{-j-1} = w_j$, $j = 0, 1, 2, \dots$, which holds for $\alpha = 2\rho$.

Table 1: Weighting patterns for future state model

ρ	α	all positive	balance
zero	$\alpha \geq 0$	yes	symmetric
positive	$0 < \rho\alpha \leq 1$	yes	past
positive	$1 < \rho\alpha < 2$	no	past
positive	$\rho\alpha = 2$	no	skew-symmetric
positive	$\rho\alpha > 2$	no	future
negative	$\alpha > 0$	yes	future

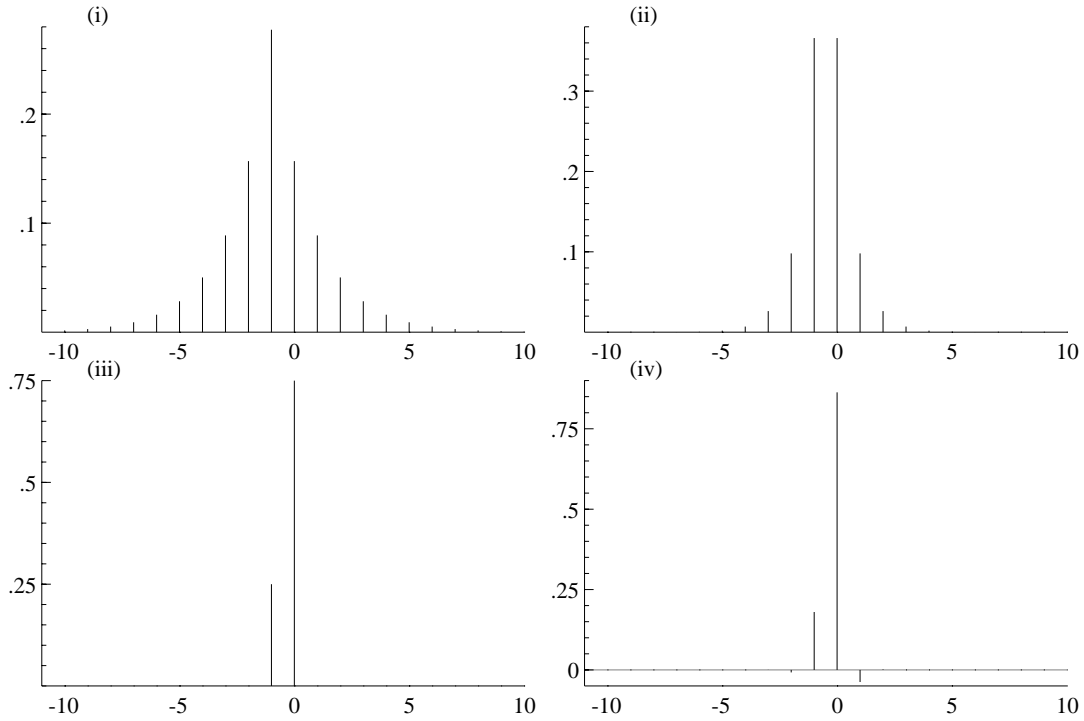


Figure 1: Weighting patterns for $\rho = .5$ and (i) $\alpha = .5$, (ii) $\alpha = 1$, (iii) $\alpha = 2$ and (iv) $\alpha = 2.5$

Applying the formula in Whittle (1983, p.58) indicates that the error in estimating the signal in a model of the form (1) has an ACGF given by

$$\{\gamma_\mu(L)\gamma_\varepsilon(L) - \gamma_{\mu\varepsilon}(L)\gamma_{\varepsilon\mu}(L)\}/\gamma_y(L). \quad (24)$$

The MSE of the estimator of the signal is given by the variance of the estimation error which is obtained as the coefficient of L^0 in (24). When the signal is a random walk

$$MSE(\tilde{\mu}_t) = \sigma_\varepsilon^2[(1 + \theta)/(1 - \theta)](1 - \rho^2) \quad (25)$$

The smoothed estimator has zero MSE when $\rho = 1$.

2.3 Perfect correlation

When $\rho = 1$ and $0 < \alpha < 2$, the invertible solution is $\theta = \alpha - 1$. Substituting $\alpha = 1 + \theta$ in (20) gives

$$w(L) = \frac{(1 + \theta)[1 + (\alpha - 1)L^{-1}]L}{|1 + \theta L|^2} = \frac{(1 + \theta)L}{1 + \theta L},$$

and so the smoothed estimator is exactly the same as the filtered estimator, (14), lagged one period, with $\lambda = \alpha$. There is a simple explanation as to why this happens. With $\rho = 1$, the model becomes

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & t = 1, \dots, T, \\ \mu_{t+1} &= \mu_t + \alpha \varepsilon_t, & \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \end{aligned} \quad (26)$$

and differencing gives

$$\Delta y_t = \varepsilon_t + (\alpha - 1)\varepsilon_{t-1}, \quad t = 2, \dots, T, \quad (27)$$

showing immediately that ξ_t is the same as ε_t and that $\theta = \alpha - 1$. Now (26) is the *innovations form* of any model of the form (1) and (2) when $\alpha < 2$ and the Kalman filter is in a steady-state; see appendix A for some details on the Kalman filter. In this case μ_t in (26) coincides with the filtered estimator, $\tilde{\mu}_{t|t-1}$. Hence it is an exact function of the past observations and so smoothing is superfluous.

In a finite sample the innovations form is

$$\begin{aligned} y_t &= \tilde{\mu}_{t|t-1} + \nu_t, & t = 1, \dots, T, \\ \tilde{\mu}_{t+1|t} &= \tilde{\mu}_{t|t-1} + k_t \nu_t, & \nu_t \sim WN(0, \sigma_t^2), \end{aligned} \quad (28)$$

where ν_t is the innovation and k_t is the Kalman gain. Provided $\alpha > 0$, σ_t^2 and k_t converge exponentially fast to σ^2 and α respectively. If the Kalman filter is started off with a diffuse prior, they do so from above; see Harvey (1989, pp.123-124) for a related discussion. A steady-state is achieved from the beginning if ε_1 is set to zero so that $\nu_t = \varepsilon_t = \xi_t$ for $t = 2, \dots, T$.

Ord, Koehler and Snyder (1997) propose a class of structural time series models based on a single disturbance term and suggest computing the likelihood function by conditioning on the initial state. In the context of (26) fixing μ_0 is not quite the same as setting ε_1 to zero and it has the disadvantage that μ_0 is unknown and so has to be estimated by a search procedure along with the parameters. However, Ord, Koehler and Snyder are concerned primarily with nonlinear models and in such cases fixing the initial state may be the best way to proceed.

If $\alpha - 1$ is regarded as a moving average parameter in (27) invertibility requires that $0 < \alpha < 2$. Within this range, a value of α less than one corresponds to the usual EWMA interpretation with the weights in (14) all positive; a value greater than one means alternate positive and negative weights.

When $\alpha = 2$, the process is strictly noninvertible. However, computing the weights by the general algorithm of appendix A confirms that they still show the skew-symmetric pattern noted earlier for models with $\rho\alpha = 2$. If $\alpha > 2$ the invertible moving average parameter in (5) is $\theta = 1/(\alpha - 1)$ and so the smoothed estimator of y_t is given by

$$w(L)y_t = \frac{(1 + \theta)^2(\alpha - 1)}{\alpha} \frac{\{1 + (\alpha - 1)^{-1}L\}}{|1 + \theta L|^2} y_t = \frac{(1 + \theta)y_t}{1 + \theta L^{-1}} = (1 + \theta) \sum_{j=0}^{\infty} (-\theta)^{|j|} y_{t+j}. \quad (29)$$

This is a *backward* filter, that is an EWMA of current and future observations. The weights alternate in sign as θ is positive.

When $\rho = -1$, so there is perfect negative correlation, differencing (26) gives

$$\Delta y_t = \varepsilon_t - (\alpha + 1)\varepsilon_{t-1}, \quad t = 2, \dots, T,$$

so the moving average parameter is $-(\alpha + 1) \leq -1$ which is noninvertible. The invertible solution is $\theta = -1/(\alpha + 1)$, which means that $-1 \leq \theta \leq 0$. The smoothed estimator is exactly as in (29), namely

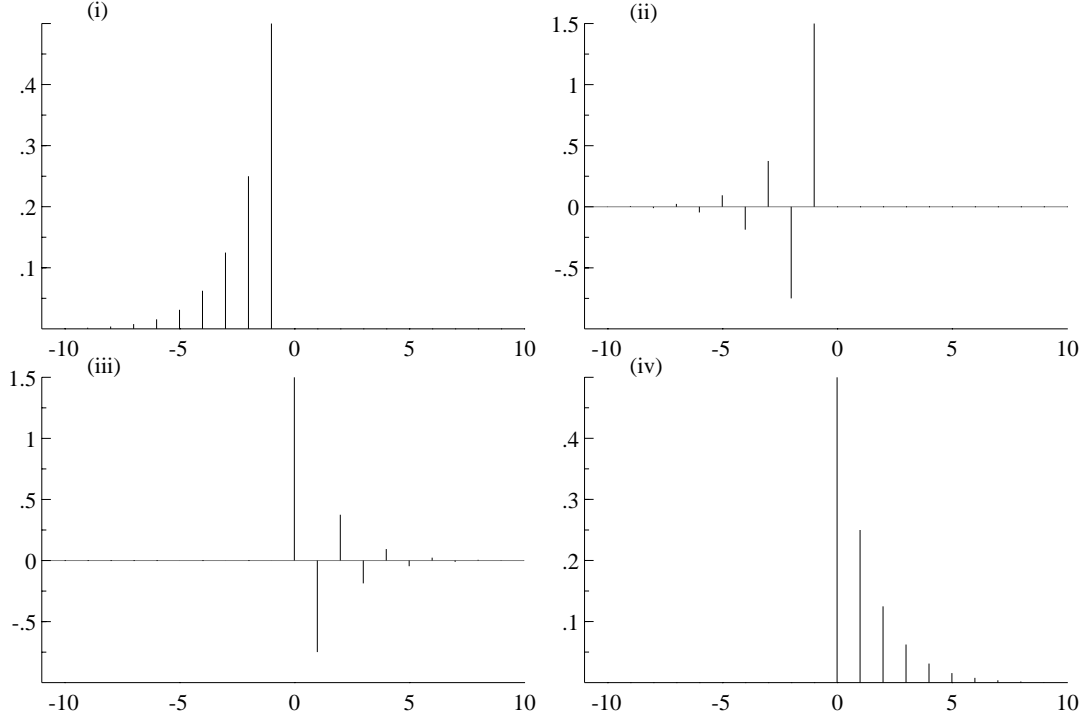


Figure 2: Weighting patterns for (i) $\rho = 1$ and $\alpha = .5$, (ii) $\rho = 1$ and $\alpha = 1.5$, (iii) $\rho = 1$ and $\alpha = 3$ and (iv) $\rho = -1$ and $\alpha = 1$

a backward EWMA. Note that a model with $\rho = 1$ but $\alpha > 2$ when $\rho = 1$ implies a reduced form in which the parameter covers the full invertibility region. However, there is no longer the direct link with the innovations form (28).

Figure 2 illustrates the above weighting patterns.

2.4 Signal extraction in the contemporaneous state model

The results in sub-section 2.2 are modified in a rather interesting way if the transition equation (2) is replaced by (4). The weights for the smoother are given by

$$w(L) = \frac{\sigma_\varepsilon^2}{\sigma^2} \frac{\alpha^2 + \rho\alpha(1 - L^{-1})}{|1 + \theta L|^2} = \frac{-\alpha\theta}{\rho\alpha + 1} \frac{(\alpha + \rho - \rho L^{-1})}{|1 + \theta L|^2} = \frac{(1 + \theta)^2}{\alpha} \frac{(\alpha + \rho - \rho L^{-1})}{|1 + \theta L|^2},$$

and so

$$\tilde{\mu}_t = \frac{1 + \theta}{1 - \theta} \left[\left(1 + \frac{\rho}{\alpha}\right) \sum_{-\infty}^{\infty} (-\theta)^{|j|} y_{t+j} - \frac{\rho}{\alpha} \sum_{-\infty}^{\infty} (-\theta)^{|j|} y_{t+j+1} \right]. \quad (30)$$

The MSE is as in (25).

If the sign of ρ is changed, these weights are a mirror image of those obtained for (2). This can be explained by the fact that if the direction of time in (4) is reversed we get

$$y_s = \mu_s + \varepsilon_s, \quad \mu_{s+1} = \mu_s - \eta_s, \quad s = 1, \dots, T,$$

which corresponds to (1) and (2). Now when ρ is positive, θ is negative and past observations receive relatively more weight.

Table 2 summarizes these findings.

Table 2: Weighting patterns for contemporaneous state model

ρ	α	all positive	balance
zero	$\alpha \geq 0$	yes	symmetric
negative	$0 < \rho\alpha \leq 1$	yes	future
negative	$1 < \rho\alpha < 2$	no	future
negative	$\rho\alpha = 2$	no	skew-symmetric
negative	$\rho\alpha > 2$	no	past
positive	$\alpha > 0$	yes	past

2.5 Perfect correlation and the Beveridge-Nelson decomposition

Now consider the case of perfect correlation for the contemporaneous state model. With $\rho = 1$, the invertible MA parameter is $\theta = -1/(\alpha + 1)$. When $\rho = -1$, the invertible MA parameter is $\theta = 1/(\alpha - 1)$ for $\alpha > 2$ and $\theta = \alpha - 1$ for $\alpha < 2$. These relationships are exactly the same as for the future state model, (2), with the sign of ρ changed.

With $\rho = 1$, the smoothed estimator is the same as the filtered estimator. The same is true when $\rho = -1$ and $\alpha > 2$, but in this case there are alternating negative and positive weights. The smoother is a backward filter when ρ is negative and $\alpha < 2$.

Beveridge and Nelson (1981) proposed that an ARIMA(0,1,1) model with $|\theta| < 1$ be decomposed into random walk and irregular components as in model (1) and (4) with

$$\eta_t = (1 + \theta)\xi_t, \quad \varepsilon_t = -\theta\xi_t. \quad (31)$$

If $\rho = 1$, the differenced UC model can be written as

$$\Delta y_t = (\alpha + 1)\varepsilon_t - \varepsilon_{t-1} = \xi_t + \theta\xi_{t-1}, \quad t = 2, \dots, T,$$

where $\theta = -1/(\alpha + 1)$ and $\xi_t = (\alpha + 1)\varepsilon_t = -\varepsilon_t/\theta$. Applying the Beveridge-Nelson (BN) decomposition as in (31) gives $\eta_t = (1 + \theta)\xi_t = -\alpha\theta\xi_t = \alpha\varepsilon_t$, so the disturbances are exactly as in the UC model. This correspondence only holds for $\theta \leq 0$. For positive θ the BN decomposition corresponds to the UC model with $\rho = -1$ and $\alpha > 2$.

The BN decomposition is defined by the requirement that the random walk component be equal to the long term prediction of the level. Hence, as noted in Watson (1986, p.55) and Harvey (1989, pp.288-289), it has zero MSE. This is consistent with (16). When $\rho = 1$, $\alpha = -(1 + \theta)/\theta$ and so $MSE(\tilde{\mu}_{t|t}) = 0$. Similarly when $\rho = -1$ and $\alpha > 2$, $\alpha = (1 + \theta)/\theta$ and again $MSE(\tilde{\mu}_{t|t}) = 0$. On the other hand, when $\rho = -1$ and $\alpha < 2$, $\theta = \alpha - 1$ and so

$$MSE(\tilde{\mu}_{t|t}) = \sigma^2(1 - \theta^2).$$

2.6 Relationship between forward and contemporaneous forms

When $\rho = 0$ both (2) and (4) lead to the same formula for the smoothed estimator of the level. However, although the models are observationally equivalent, they are not the same. The nature of the difference becomes clearer when we examine the relationship between the two models with any value of ρ .

In order to compare the models, it is helpful to elaborate the notation slightly by writing (1) and (2) as

$$\begin{aligned} y_t &= \mu_t^* + \varepsilon_t^*, & \varepsilon_t^* &\sim WN(0, \sigma_\varepsilon^{*2}), & t = 1, \dots, T, \\ \mu_{t+1}^* &= \mu_t^* + \eta_t, & \eta_t &\sim WN(0, \sigma_\eta^2), \end{aligned} \quad (32)$$

with $\alpha^* = \sigma_\eta / \sigma_\varepsilon^*$ and $Cov(\varepsilon_t^* \eta_t) = \rho^* \alpha^* \sigma_\varepsilon^{*2}$. The notation for the model based on (4) will remain the same. The above future state form model can be written in contemporaneous state form simply by letting $\mu_t^* = \mu_{t-1}$. The measurement equation is then

$$y_t = \mu_{t-1} + \varepsilon_t^* = \mu_t + \{\varepsilon_t^* - \eta_t\}. \quad (33)$$

Now denote the weighting function in (21) as $w^*(L)$, so that $\tilde{\mu}_t^* = w^*(L)y_t$. If $\tilde{\mu}_t = w(L)y_t$ then by definition $w(L) = w^*(L)L^{-1}$. This can be checked by applying the signal extraction formula for the contemporaneous state with

$$\sigma_\varepsilon^2 = \sigma_\varepsilon^{*2} + \sigma_\eta^2 + 2\rho^* \sigma_\eta \sigma_\varepsilon^*, \quad \alpha = \sigma_\eta / \sigma_\varepsilon$$

and

$$\rho = \frac{\rho^* - \alpha^*}{\sqrt{1 + \alpha^{*2} - 2\rho^* \alpha^*}} = \frac{\rho^* \alpha}{\alpha^*} - \alpha.$$

Substituting

$$\frac{\rho}{\alpha} = \frac{\rho^*}{\alpha^*} - 1. \quad (34)$$

in (30) gives (21) multiplied by L^{-1} .

Thus when ρ^* is zero, the implied value of ρ is $-\alpha$. As a result, the smoother applied to the contemporaneous form gives a weighting pattern for μ_t which, as can be seen from (30), is symmetric around $t+1$ not t . Conversely, if ρ is zero in the contemporaneous form, then $\rho^* = \alpha^*$ and the smoother for μ_t^* is symmetric around $t-1$.

2.7 Random walk

If a series follows a random walk it can be decomposed into a random walk plus noise if the correlation between the disturbances is allowed to be non-zero. In the case of (2) this requires, in the amended notation of the previous sub-section, $\rho^* \alpha^* = 1$. If we set $\rho^* = 1$, differencing the UC model gives $\Delta y_t = \varepsilon_t$, showing that ε_t is the reduced form disturbance, ξ_t . Thus $\tilde{\varepsilon}_t = \Delta y_t$ and it can be seen from (15) that the filtered estimator of μ_t is

$$\tilde{\mu}_{t|t} = y_t - \Delta y_t = y_{t-1}.$$

The smoothed estimator is the same (as is $\tilde{\mu}_{t|t-1}$). Both have zero MSE.

For (4), $\rho\alpha = -1$ and if $\rho = -1$ the smoothed estimator of μ_t is y_{t+1} . The filtered estimator is y_t . Unlike the smoothed estimator, which is equivalent to a backward filter, the filtered estimator has a MSE of σ^2 rather than a MSE of zero. This can be seen to be the case because

$$y_{t+1} = \mu_{t+1} + \varepsilon_{t+1} = \mu_t - \varepsilon_{t+1} + \varepsilon_{t+1} = \mu_t.$$

and so $y_{t+1} - \tilde{y}_{t+1|t} = \mu_t - \tilde{\mu}_{t+1|t} = \mu_t - \tilde{\mu}_{t|t}$. Hence the MSE of $\tilde{\mu}_{t|t}$ is the same as that of $\tilde{y}_{t+1|t}$.

These rather strange results lend some force to the view that it makes little sense to try to decompose a random walk into stationary and nonstationary components.

2.8 Continuous time and temporal aggregation

Other things being equal there seems to be little reason for formulating a model with correlated disturbances. However, suppose a model is set up in continuous time with uncorrelated incremental disturbances, but that the process, $y(t)$, $0 \leq t \leq T$, is observed at discrete intervals as an integral, that is

$$y_t = \int_{t-1}^t y(s)ds, \quad t = 1, \dots, T.$$

The discrete time model is then of the form (32) with correlated disturbances; see Harvey (1989, p.495). The correlation is such that

$$\rho^* = \alpha^*/2, \quad (35)$$

or, if the model is cast in contemporaneous form, $\rho = -\alpha/2$. Either way the result is a smoother which, for μ_t , is symmetric around $t + 1/2$, that is $w_{-j} = w_{j+1}$, $j = 0, 1, 2, \dots$. More specifically, $w_0 = w_1 = (1 + \theta)/2$ with the weights declining exponentially at either side. This is what one would expect since y_t contains information on the continuous time level accumulated between $t - 1$ and t , while y_{t+1} contains information between t and $t + 1$.

As well as providing an instance where there is a theoretical rationale for correlated disturbances, the continuous time model also allows a reduced form with positive θ , a value for θ of 0.27 arising in the extreme case when $y(t)$ is pure Brownian motion; see Harvey (1989, p.496).

A discrete time model formulated at a unit time interval but observed every δ time periods provides an intermediate case. From Harvey (1989, p.317), it can be deduced that the correlation between the disturbances is $\rho^* = \alpha^*(\delta + 1)/2\delta$, $\delta = 1, 2, \dots$

3 Local linear trend model

The local linear trend model contains a stochastic slope. Thus, we have

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim WN(0, \sigma_\varepsilon^2), \\ \mu_{t+1} &= \mu_t + \beta_t + \eta_t, & \eta_t &\sim WN(0, \alpha^2 \sigma_\varepsilon^2), \\ \beta_{t+1} &= \beta_t + \zeta_t, & \zeta_t &\sim WN(0, \gamma^2 \sigma_\varepsilon^2). \end{aligned} \quad t = 1, \dots, T, \quad (36)$$

Correlated disturbances would lead to asymmetric weighting patterns the form of which could be explored as in section 2. Perfect correlation leads to a model which is effectively in innovations form and restrictions on α and γ are needed to ensure invertibility of the reduced form. There is also the issue of whether patterns such as the alternating positive and negative weights should be avoided. We will not discuss these matters in detail as there are no new substantive issues involved. Instead we concentrate on the weighting pattern when the three disturbances are mutually uncorrelated.

Figure 3 shows the weights from the KFS with various values of α and γ . It is interesting that there can be some negative weights; this could conceivably result in a negative trend at certain points even though all the observations are positive. The weights for the smoothed estimator of the slope at time t are given simply by subtracting the weights for the level at time $t + 1$ from those at time t .

When $\alpha = 0$ the WK formula gives

$$w(L) = \frac{\sigma_\zeta^2}{\sigma^2 |1 + \theta_1 L + \theta_2 L^2|^2}, \quad (37)$$

where $\sigma_\zeta^2 > 0$ is the variance of ζ_t . (If $\sigma_\varepsilon^2 = 0$, the expression is still valid with $w_0 = 1$ and the other weights zero). By equating the autocovariances at lags one and two it can be shown that the reduced form parameters satisfy $\theta_1 = -4\theta_2/(1 + \theta_2)$, with $0 \leq \theta_2 < 1$ and $-2 < \theta_1 \leq 0$. The roots of the polynomial $1 + \theta_1 L + \theta_2 L^2$ are complex. Since $1/|1 + \theta_1 L + \theta_2 L^2|^2$ is the ACGF of an AR(2) process,

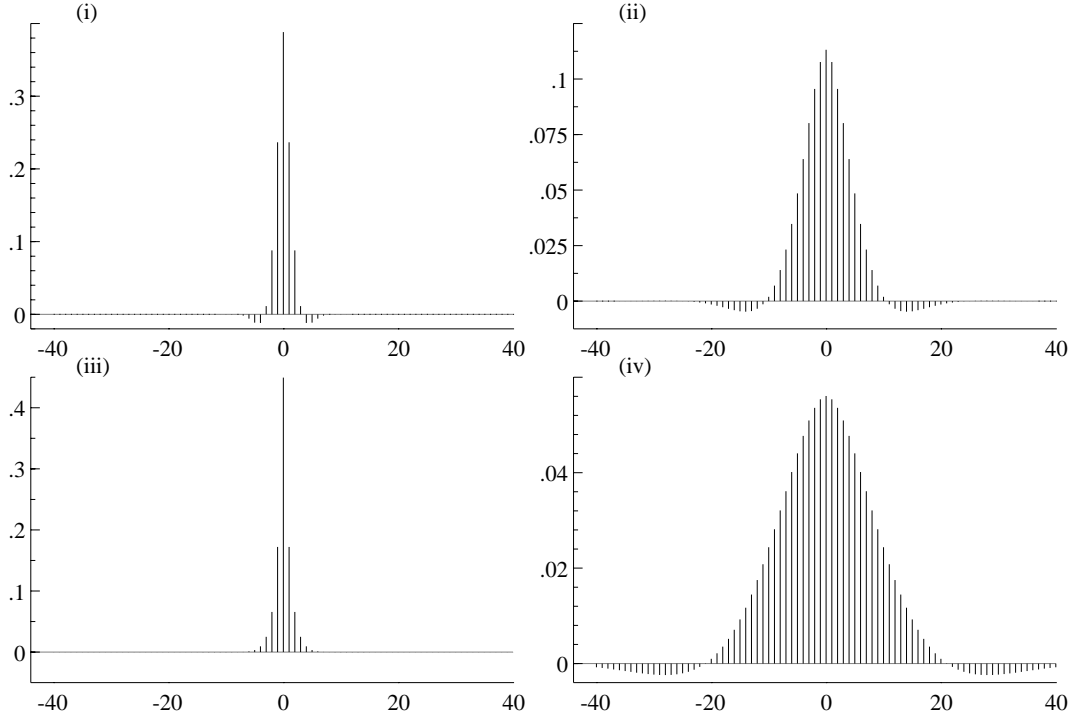


Figure 3: Weighting patterns for (i) $\alpha = 0$ and $\gamma = 1$, (ii) $\alpha = 0$ and $\gamma = 0.1$, (iii) $\alpha = 1$ and $\gamma = 0.1$ and (iv) $\alpha = 0$ and $\gamma = 0.025$

it follows that the weights decay according to a damped sine wave with frequency $\cos^{-1}[-\theta_1/2\sqrt{\theta_2}]$; see Box and Jenkins (1976, p.59). Hence the negative weights observed in the first three panels of figure 3. The value of $\gamma = 0.025$ corresponds to the filter proposed by Hodrick and Prescott (1997) for quarterly observations; see Harvey and Jaeger (1993).

In the more general case when α is not necessarily zero, the weighting function is of the form

$$w(L) = \frac{\sigma_\mu^2}{\sigma^2} \frac{|1 + \theta_\mu L|^2}{|1 + \theta_1 L + \theta_2 L^2|^2}, \quad (38)$$

where the numerator is obtained from the reduced form of the trend component, which is such that $\Delta^2 \mu_t$ is an $MA(1)$ with MA parameter θ_μ and disturbance variance σ_μ^2 . The weights are now obtained from the ACGF of an $ARMA(2, 1)$ process in which $-1 \leq \theta_\mu \leq 0$ and, as before, $0 \leq \theta_2 < 1$ and $-2 < \theta_1 \leq 0$. The roots of the implied autoregressive polynomial need no longer be complex, but if they are, the damped sine wave starts from $j = 1$ rather than zero. If the roots are real they must be non-negative. Note that the structural model with mutually uncorrelated disturbances implies that only a small part of the invertibility region of the reduced form $ARIMA(0, 2, 2)$ parameter space is admissible; see Harvey (1989, p.69).

The model may be extended so that it becomes a local approximation to a quadratic, or indeed any polynomial. For a polynomial of order m the reduced form will be $ARIMA(0, m + 1, m + 1)$ and weighting patterns may, in principle, be derived from the WK formula. Seasonal and cyclical components may also be added, leading to more complex weighting patterns; Riani (1998) investigates some of these patterns numerically.

4 Signal Extraction from ARIMA Models

The discussion so far has assumed the model is formulated directly in UC form. However, decompositions are possible for ARIMA models. The Beveridge-Nelson filter, introduced in sub-section 2.5, can be computed for any ARIMA model which is stationary and invertible in first differences, but it is one-sided and so may not be appealing for signal extraction. The first sub-section below gives a corresponding symmetric two-sided filter. The second sub-section reviews the canonical decomposition.

4.1 A Beveridge-Nelson signal extraction filter

Consider an ARIMA($p, 1, q$) model

$$\Delta y_t = [\theta(L)/\phi(L)]\xi_t, \quad t = 2, \dots, T, \quad (39)$$

where $\theta(L)$ and $\phi(L)$ are polynomials in the lag operator and ξ_t is a white noise disturbance with variance σ^2 . The BN filter is defined to be the value of the forecast in the limit as the lead time goes to infinity and it yields a trend given by

$$\mu_t^{BN} = \frac{\theta(1)}{\phi(1)} \frac{\phi(L)}{\theta(L)} y_t = w^{BN}(L) y_t \quad (40)$$

Thus it is a function of current and past observations. Proietti and Harvey (1999) suggest that a corresponding two-sided filter, the BN smoother, be constructed as:

$$\mu_t^{BNS} = \left| w^{BN}(L) \right|^2 y_t \quad (41)$$

Now consider a UC model made up of two uncorrelated components, a random walk and a stationary ARMA(p, q^*) process. Some ARIMA($p, 1, q$) models can be decomposed in this way and in such cases the optimal smoother is given by the BN smoother of (41). If there is no such decomposition the appeal of the BN smoother, and perhaps also of the BN filter itself, may be limited.

When $\theta \leq 0$, an ARIMA(0, 1, 1) model, as in (5), can be decomposed into random walk and noise components driven by disturbances which are uncorrelated with each other. It is easy to see that the BN smoother is the same as the smoother in this UC model. When θ is positive there is no such decomposition and the BN filter has alternating negative and positive weights.

The BN trend in the ARIMA(1, 1, 0) model,

$$\Delta y_t = \phi \Delta y_{t-1} + \xi_t, \quad t = 2, \dots, T,$$

is

$$\mu_t^{BN} = y_t + \frac{\phi}{1-\phi} \Delta y_t = \frac{1}{1-\phi} y_t - \frac{\phi}{1-\phi} y_{t-1}, \quad t = 2, \dots, T.$$

The corresponding smoother is therefore

$$\mu_{t|T}^{BNS} = \frac{-\phi}{(1-\phi)^2} y_{t-1} + \frac{1+\phi^2}{(1-\phi)^2} y_t - \frac{\phi}{(1-\phi)^2} y_{t+1}, \quad t = 2, \dots, T-1.$$

If $\phi \leq 0$ the model decomposes into uncorrelated random walk and AR(1) components with the variance of the disturbance in the AR(1) being equal to $-\phi\sigma^2/(1-\phi)^2$. In this case the weights in the BN smoother are all positive.

4.2 Canonical Decomposition

The canonical decomposition can be applied to any ARIMA(p, d, q) model provided that $q \geq p + d + 1$. The ARIMA(0, 1, 1) model is a valid reduced form for a UC model of the form

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \quad t = 1, \dots, T$$

$$\mu_t = \mu_{t-1} + \eta_t + \theta^\dagger \eta_{t-1}, \quad \eta_t \sim WN(0, \alpha^2 \sigma_\varepsilon^2) \quad (42)$$

Assume $E(\varepsilon_s \eta_t) = 0$ for all s, t . As with correlated disturbances there is an identifiability problem, and the usual solution is for θ^\dagger to be set to a given value. In the canonical decomposition of Hillmer and Tiao (1982) θ^\dagger is set to one, thereby maximising the variance of the noise, ε_t . (Although this UC representation is normally constructed from an estimated ARIMA(0, 1, 1) model, there is no reason why it should not be estimated directly.)

The application of the WK signal extraction formula gives

$$w(L) = \frac{\sigma_\varepsilon^2 \alpha^2 |1 + \theta^\dagger L|^2}{\sigma_\varepsilon^2 |1 + \theta L|^2} = \frac{(1 + \theta)^2 |1 + \theta^\dagger L|^2}{(1 + \theta^\dagger)^2 |1 + \theta L|^2}, \quad (43)$$

the second step following from expressions for the variance and first-order autocovariance. This yields

$$\begin{aligned} w_0 &= [(1 + \theta)/(1 - \theta)(1 + \theta^\dagger)^2][1 + \theta^{\dagger 2} - 2\theta\theta^\dagger], \\ w_j &= [(1 + \theta)/(1 - \theta)(1 + \theta^\dagger)^2][\theta^\dagger - \theta\theta^{\dagger 2} - \theta + \theta^2\theta^\dagger](-\theta)^{|j-1|}, \quad j = 1, 2, \dots \end{aligned}$$

Thus with $\theta^\dagger > 0$ (< 0) the weights start their exponential decline at $j = 1$ at a point relatively higher (lower) than if $\theta^\dagger = 0$. This may or may not be thought to be desirable. However, two points should be noted. The first is that when the observations follow a random walk, the smoothing weights are $w_0 = 1/2$ and $w_1 = w_{-1} = 1/4$. The second is that unless $\theta^\dagger = 0$, the filtered estimator at the end of the series, $\tilde{\mu}_{T|T}$, will not be equal to the eventual forecast function, $\tilde{\mu}_{T+\ell|T}$, $\ell \geq 1$.

5 Heteroscedasticity and Irregular Spacing

Heteroscedasticity arising from a slowly changing variance in ε_t may be handled by constructing a stochastic model. One possibility is the stochastic volatility model in which $Var(\varepsilon_t) = \sigma_\varepsilon^2 \exp(h_t)$ where h_t is a stochastic process such as AR(1), random walk or local linear trend; see Ghysels *et al* (1996). Other possibilities include letting σ_ε^2 or σ_ε evolve according to a stochastic process. ML methods for fitting such models are beyond the scope of this paper. Suffice to say that simulation methods will need to be used; see Shephard (1994) and Sandmann and Koopman (1998). A simple method is described in connection with the illustration in section 8. The point is that heteroscedasticity means that the signal-noise ratio changes throughout the series, but the signal extraction weights adapt accordingly.

Irregular spacing of observations poses no problem if the model is formulated in continuous time; see Harvey (1989, chapter 9). For a local level model with uncorrelated disturbances in continuous time, the implied discrete time model for a stock variable observed at times $\tau_t, t = 1, \dots, T$ is as in (1),(2), but with $Var(\eta_t) = \sigma_\varepsilon^2 \alpha^2 \delta_t$, where $\delta_t = \tau_{t+1} - \tau_t$ is the time between observation t and observation $t + 1$. (Note that with a flow variable the implied discrete time model is as in sub-section 2.8.) It is possible to have two (or more) observations located at the same point; thus if the t -th and $t + 1$ -th observations are observed at the same time, $\delta_t = 0$. Application of the KFS gives an optimal weighting pattern. This may be computed explicitly using the method outlined in appendix A; the derivation can be found in Koopman and Harvey (1999).

6 Robust Signal Extraction

The signal extraction formulae of section 2 are optimal in the sense that they minimise the MSE within the class of linear estimators. If the disturbances are Gaussian, they are optimal within the class of all estimators. We now consider the implications for weighting patterns when the disturbances are specified as having non-Gaussian distributions. The main motivation for doing this is to make prediction and signal extraction robust to outliers and structural breaks.

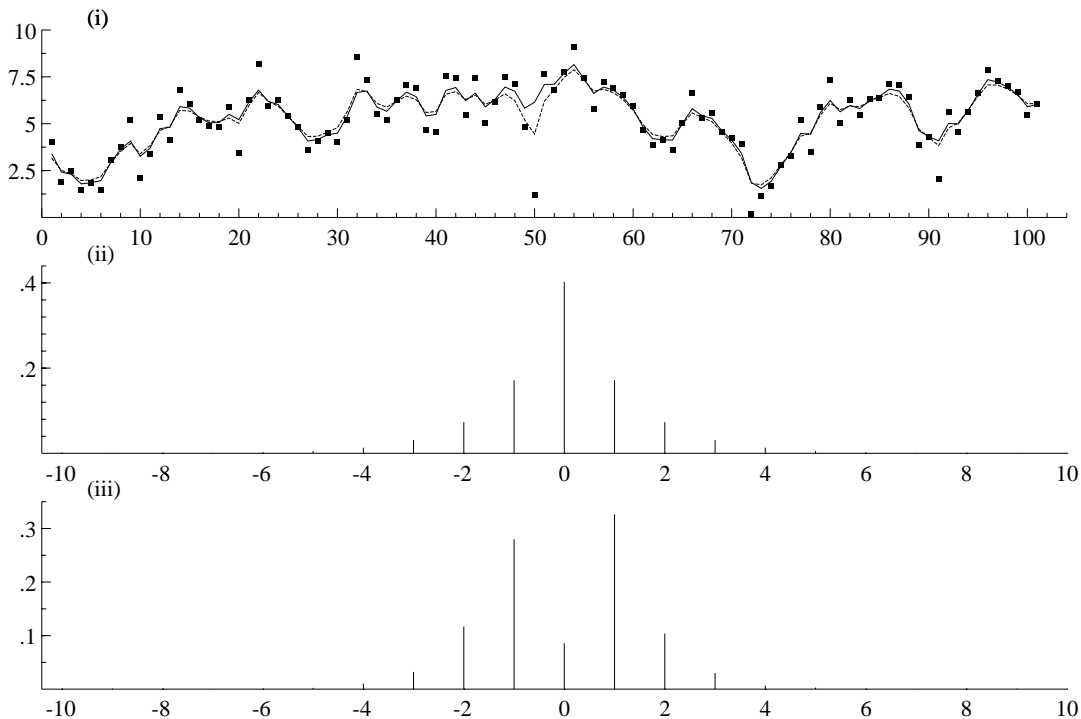


Figure 4: Data with outlier. (i) simulated data, Gaussian signal (dotted) and t signal (solid), (ii) weighting pattern for Gaussian signal $t = 50$ and (iii) weighting pattern for t signal.

Outliers can produce distortions in the estimated signal. Figure 4 (i) shows an artificial data set generated by a Gaussian local level model with $\sigma_\varepsilon^2 = 1$ and $\sigma_\eta^2 = 0.64$ and with an outlier at $t = 50$ formed by multiplying the original disturbance by a factor of 3.5. Estimating a Gaussian model gives $\tilde{\sigma}_\varepsilon^2 = 1.013$ and $\tilde{\sigma}_\eta^2 = 0.781$ and it can be seen how the outlying observation adversely affects the estimated signal. It should therefore be given less weight. We could treat it as missing, which is effectively the same as letting the variance of the measurement error go to infinity. This strategy is perhaps too extreme since the observation may still contain information. A better option is to multiply the measurement variance by some factor greater than one. Since prior information for setting such factors is rarely available, we set up a model in which the measurement error is assumed to come from a distribution with long tails. Durbin and Koopman (1999) show that, in the case of a local level model with t -distributed measurement errors, the mode of the posterior distribution of μ_t can be computed by a single KFS applied to a Gaussian model in which

$$\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2 h_t\right), \quad t = 1, \dots, T.$$

This allows an analysis of weighting patterns implied by the robust non-Gaussian model. The h_t 's are

obtained as follows. We define

$$h_t = \frac{1}{\nu + 1} \tilde{\varepsilon}_t^2 + \frac{\nu - 2}{\nu + 1} \sigma_\varepsilon^2, \quad t = 1, \dots, T.$$

where σ_ε^2 and ν are the variance and degrees of freedom respectively of the t -distribution, and $\tilde{\varepsilon}_t = y_t - \tilde{\mu}_t$ is computed by the KFS applied to an initial Gaussian model with $h_t = 1$. The KFS applied to the model with measurement variances $\sigma_\varepsilon^2 h_t$ will produce new $\tilde{\varepsilon}_t$'s and new h_t 's can be computed from these. The number of iterations required for a reasonable level of convergence is usually small, that is, around ten.

The above method requires that the parameters σ_ε^2 , σ_η^2 and ν are known. Durbin and Koopman (1997) show how they can be estimated by Monte Carlo maximum likelihood (MCML); Shephard and Pitt (1997) give the details of how they can be estimated by Bayesian methods. Using MCML, the degrees of freedom of the t -distribution was estimated to be 3.5 while $\tilde{\sigma}_\varepsilon^2 = 1.143$ and $\tilde{\sigma}_\eta^2 = .735$. The resulting t signal, shown in figure 4 (i), is not distorted by the outlier. The implied weight patterns for the Gaussian and t signals at $t = 50$ are presented in figure 4, and it can be seen how the outlying observations are downweighted for the t signal. Since the model is not time invariant the classical WK formula cannot be used and the weights were computed by method set out in appendix A.

We may also wish to estimate signals which are robust to structural breaks. To allow for such breaks, the level disturbance, η_t , may be assumed to have a t -distribution. The procedure works in much the same fashion but now we introduce the multiplication factor q_t such that

$$\eta_t \sim N(0, \sigma_\eta^2 q_t), \quad t = 1, \dots, T.$$

The q_t 's are obtained by a sequence of KFS runs and for each run we have

$$q_t = \frac{1}{\nu + 1} \tilde{\eta}_t^2 + \frac{\nu - 2}{\nu + 1} \sigma_\eta^2, \quad t = 1, \dots, T,$$

where $\tilde{\eta}_t = \tilde{\mu}_{t+1} - \tilde{\mu}_t$.

Figure 5 (i) shows a series generated by a Gaussian model with $\sigma_\varepsilon^2 = 1$ and $\sigma_\eta^2 = 0.64$ and the original η_{50} multiplied by 3.5, thereby creating a change in the level. The estimates for a Gaussian model are $\tilde{\sigma}_\varepsilon^2 = 1.186$ and $\tilde{\sigma}_\eta^2 = .843$. As can be seen, the estimated signal does not respond immediately to the break, and the price paid for attempting to do so is that it is too variable in the rest of the series because of the higher signal-noise ratio of .71. A t -distribution fitted to the level disturbance has degrees of freedom 4.2 and $\tilde{\sigma}_\varepsilon^2 = 1.365$ and $\tilde{\sigma}_\eta^2 = .724$. As can be seen in figure 5, the weights for $\tilde{\mu}_{50}$ are larger for the data-points $t = 50, 49, 48$ as compared with $t = 51, 52, 53$ while for $\tilde{\mu}_{51}$ the converse is true. Thus a locally asymmetric weight pattern is used to produce a robust signal.

7 Nonparametric Signal Extraction

There is a very close link between the so called nonparametric techniques for signal extraction and those based on UC time series models.

7.1 Kernel estimation

If the trend in a time series is regarded as a deterministic function of unspecified form, it may be estimated at all points by a weighted moving average, the shape of this moving average being termed a kernel. By adopting the rather artificial device that observations are assumed to arrive more frequently (in a given time interval) as the sample size increases, it can be shown that a suitably designed kernel will estimate the trend consistently ; see, for example, Härdle (1990). The method assumes that the

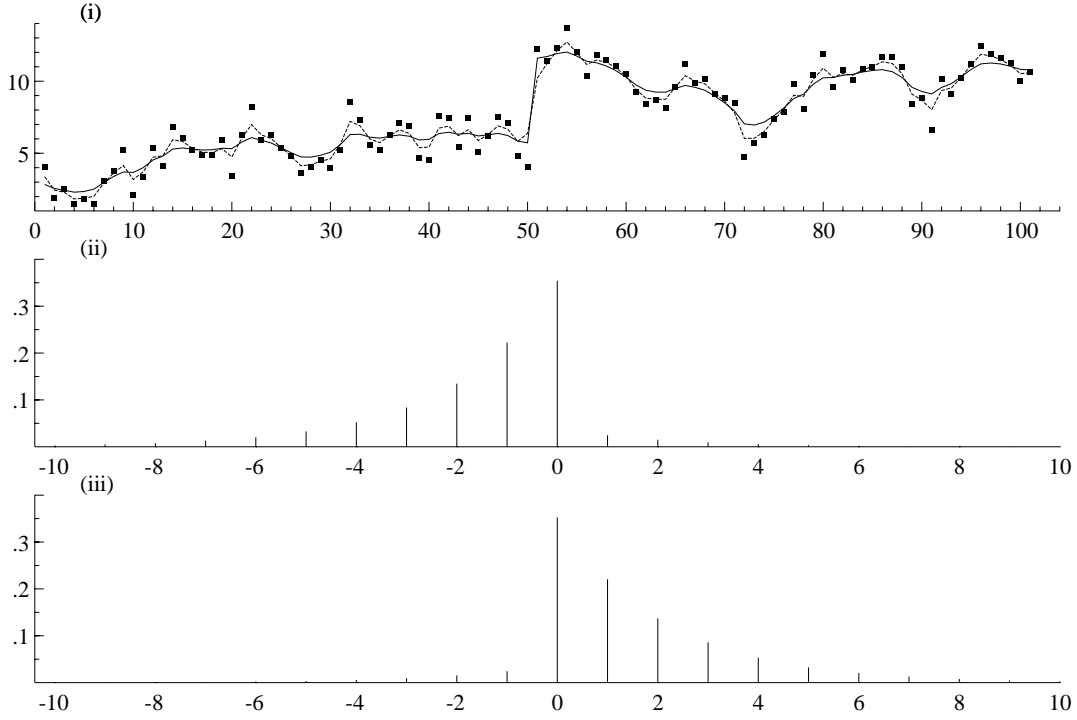


Figure 5: Data with break. (i) simulated data, Gaussian signal (dotted) and t signal (solid), (ii) weighting pattern for t signal at $t = 50$ and (iii) weighting pattern for t signal at $t = 51$.

non-trending part of the series is white noise, but it may be extended by adopting a hybrid procedure based on fitting an ARMA model; see Hart (1994).

An alternative approach is to fit a time series model and carry out signal extraction. The implied signal extraction weights from a fitted UC time series model constitute a kernel. In the random walk plus noise model, the signal-noise ratio, α^2 , plays a similar role to a kernel bandwidth. A lower α^2 corresponds to a wider bandwidth. Gijbels, Pope and Wand (1999) explore this connection for the filter, but not the smoother.

It is interesting to examine how particular nonparametric kernels for signal extraction might be approximated by UC models. As was noted earlier a unit root is needed for the weights to sum to one, and so, for regularly spaced data, we might consider generalisations of (42) in which $\Delta\mu_t$ follows an $MA(q)$ process with the MA parameters pre-set to certain values. The pattern of weights can then be obtained in terms of the MA parameters as was done for (43); the exponential decline will start at $j = q$. The reduced form is $ARIMA(0, 1, q)$ and diagnostics can be used to check if such a model is consistent with the data. A further set of kernels could be obtained for local linear trend models by a similar device. In this case there is the possibility of extending both the level and slope disturbances to become moving averages.

The above models could be modified by replacing the irregular component by a more general stationary process. Note that the weights for UC models do not cut off to zero beyond a certain point, except in certain special cases. One such case is when the stationary component is $AR(1)$ with a disturbance variance equal to minus the AR parameter times the variance of the random walk disturbance. This constraint means that there is no moving average in the reduced form, and hence no infinite lags in either filter or smoother; see end of sub-section 4.1.

7.2 Cubic splines

The connection between cubic splines and the local linear trend model has been known for many years; see Wecker and Ansley (1983). The equivalence is actually with a local linear trend formulated in continuous time with the variance of the level disturbance equal to zero. This corresponds to a discrete time model, (36), with α positive and η_t and ζ_t correlated as in (44) below. The WK filter is as in (38) since the correlation between the level and slope disturbances does not alter the fact that the second difference of the trend component is an $MA(1)$. However, this model is not too different to the model with $\alpha = 0$ in which case the WK weighting function is as in (37). In fact calculation of the weights for discrete and continuous time models with $\gamma = 0.025$ gives values which are the same to three decimal places; only when γ is greater than one do significant differences start to become apparent. An approximate expression for the implied kernel was obtained, using a completely different method, by Silverman (1984); see also Green and Silverman (1994, p.47).

The local quadratic model in continuous time leads to a quintic spline; see Kohn, Ansley and Wong (1992). The implied kernel comes from the ACGF of an ARMA(3, 2) process.

When the observations are irregularly spaced, the discrete time model implied by the underlying continuous time model is as in (36) with

$$Var(\eta_t) = \sigma_\varepsilon^2 \gamma^2 \delta_t^3 / 3, \quad Var(\zeta_t) = \sigma_\varepsilon^2 \gamma^2 \delta_t, \quad E(\eta_t \zeta_t) = \sigma_\varepsilon^2 \gamma^2 \delta_t^2 / 2, \quad (44)$$

and ε_t uncorrelated with the disturbances in the trend; see Harvey (1989, chapter 9). As in section 5, $\delta_t = \tau_{t+1} - \tau_t$ is the time between observation t and observation $t + 1$. With evenly spaced observations the δ_t 's are set to one. The fact that irregularly spaced data may be handled means that the model can be used to fit a nonlinear function to cross-sectional data as in the example below.

8 Illustration

Here we consider 133 observations of acceleration against time (measured in milliseconds) for a simulated motorcycle accident. This data set was originally analysed by Silverman (1985) and is often used as an example of nonparametric curve fitting techniques; see, for example, Härdle (1990) and Green and Silverman (1994). The observations are not equally spaced and at certain time points there are multiple observations; see figure 6. Nonparametric cubic spline and kernel smoothing techniques depend on some choice of a smoothness parameter. This is usually determined by cross validation. However, setting up a UC model for a cubic spline enables the smoothness parameter to be estimated by maximum likelihood and the spline to be computed by the KFS. The model can easily be extended, for example to include other components, and it can be compared with alternative models using standard statistical criteria.

In the first sub-section below we fit a standard cubic spline using maximum likelihood. We then show that when the data are irregularly spaced, the weights the KFS uses to construct the spline are not the same as those used in the standard nonparametric approach. In the second sub-section we consider the local level model as an alternative and we look at its performance in relation to the cubic spline model. The third sub-section fits a spline using a simple method to correct for heteroscedasticity. The calculations were carried out using the `SsfPack` 2.2 package of Koopman *et al* (1998) implemented in the `0x` programming language of Doornik (1998). The `0x` code for the computations carried out in the next section is given in appendix B.

8.1 Cubic spline

The smoothing parameter γ is estimated by maximum likelihood (assuming normally distributed disturbances) using the transformation $\gamma = \exp(\psi)$. The estimate of ψ is -3.59 with asymptotic

standard error 0.22. The estimate of γ is 0.0275 with an asymmetric 95% confidence interval of 0.018 to 0.043. Silverman (1985) estimates γ by cross-validation, but does not report its value. In any case, it is not clear how one would compute a standard error for an estimate obtained by cross-validation. The maximized log-likelihood is $l = -624.1$ and the Akaike information criterion (AIC) is $2n - 2l = 1254$ where n is the number of state elements plus the number of estimated parameters ($2 + 1$ here); see Harvey (1989, pp.80-81) for more details. Figure 6 (i) presents the cubic spline. One of the advantages of representing the cubic spline as a statistical model is that we can compute variances of our estimates and, therefore, standardised residuals. The 95% confidence intervals for the fitted spline are also given in figure 6 (i). These are based on the root mean square errors (RMSE) of the estimates of μ_t , obtained from the KFS, but without an allowance for the uncertainty arising from the estimation of γ . The standardised smoothed irregular component presented in figure 6 (ii) clearly indicates heteroscedasticity.

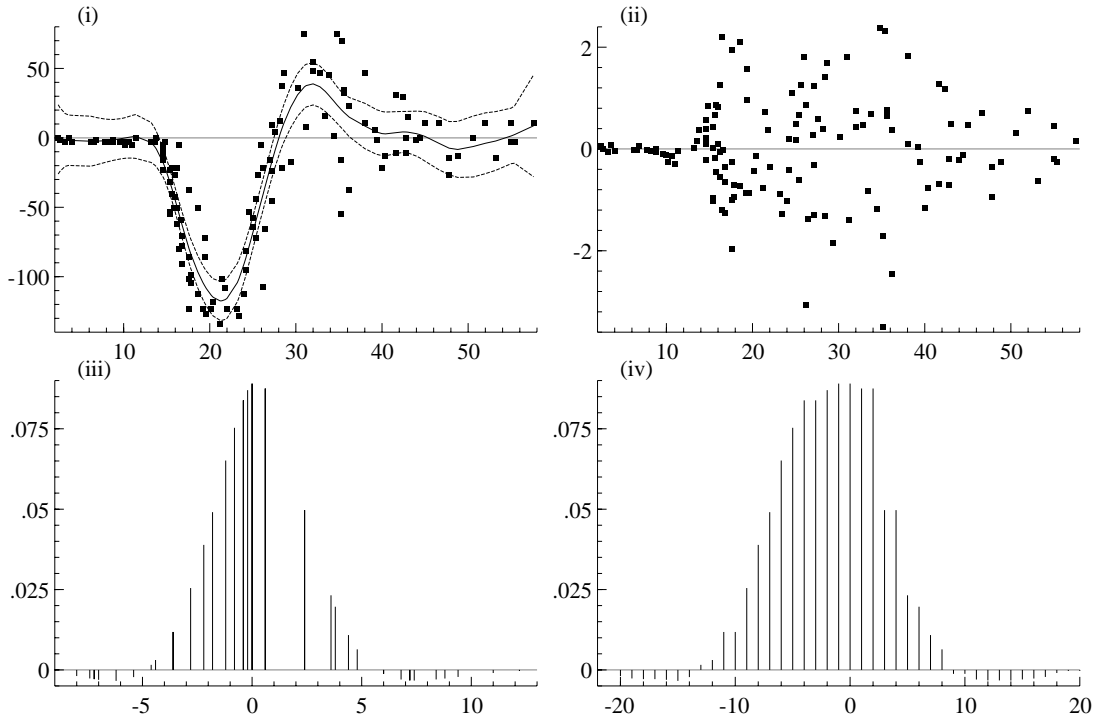


Figure 6: Motorcycle acceleration data analysed by a cubic spline. (i) observations against time with spline and confidence intervals, (ii) standardised irregular, (iii) the weights for the spline at time $\tau_{105} = 35.6$ against time distance $\tau_{105} - \tau_j$ and (iv) weights for the spline at time $\tau_{105} = 35.6$ for adjacent observations.

The motorcycle acceleration data are unequally spaced and at certain time points there are multiple observations. Weighting patterns for the smoothed estimates of the spline, μ_{τ_t} , can be presented in two ways, depending on whether they are plotted against time (or distance) from τ_t or against adjacent observations, that is observation number minus t . Figure 6 (iii) gives the weights for $\tau_t = 35.6$, where $t = 105$, on the time scale, while figure 6 (iv) presents weights for adjacent observations. Plotting against time shows only one weight for multiple observations, whereas the same weight is repeated when plotting against observation. The weighting pattern in figure 6 (iii) is *not* symmetric. This is in contrast to the nonparametric approach where the weighting pattern is symmetric in that observations

which are at the same ‘distance’ from the time point τ_i receive the same weight; see, for example, Green and Silverman (1994). The reason the optimal weights, obtained from the model, are not symmetric is that the number of data points observed around a particular observation is taken into account. An observation at a time point where relatively many observations are concentrated receives relatively less weight because it has less impact. On the other hand, when observations are relatively sparse they receive more weight.

8.2 Local level

A local level model may also be fitted and compared with the cubic spline using statistical criteria.

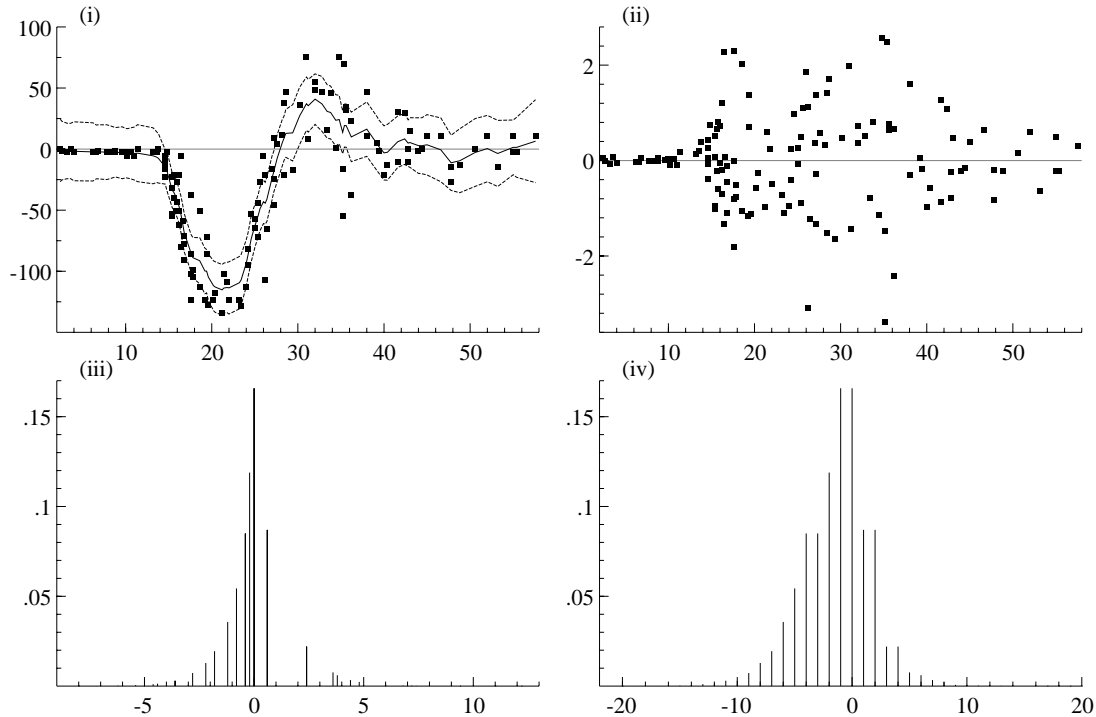


Figure 7: Motorcycle acceleration data analysed by a local level model. (i) observations against time with spline and confidence intervals, (ii) standardised irregular, (iii) the weights for the spline at time $\tau_{105} = 35.6$ against time distance $\tau_{105} - \tau_j$ and (iv) weights for the spline at time $\tau_{105} = 35.6$ for adjacent observations.

The ML estimate of the signal-noise ratio, α , is 0.33 and the maximized log-likelihood is $l = -625.9$. The AIC is 1256 (with $n = 2$) which is slightly higher than the AIC for the cubic spline model, indicating a preference for the latter.

Figure 7 presents graphs corresponding to spline graphs in figure 6. As might be expected, the signal for the local level model is less smooth than for the spline. The asymmetry in the weights against time is even more pronounced.

8.3 Heteroscedasticity

In figure 6 (ii) the standardised irregular was presented for the cubic spline model and we concluded that the errors were heteroscedastic. Rather than attempting to fit a stochastic volatility model as

outlined in section 5, we simply correct for heteroscedasticity by fitting a local level signal through the absolute values of the smoothed estimates of the irregular component; compare the suggestions for weighted nonparametric estimation in Silverman (1985). Subsequently we replace the measurement error variance, σ_ε^2 , of the original cubic spline model by $\sigma_\varepsilon^2 h_t^{*2}$ where h_t^* is the smoothed estimate of the local level signal, scaled so that $h_1^* = 1$. The h_t^* 's are always positive because the weights of a local level model with uncorrelated disturbances are always positive, see (23). The absolute values of the smoothed irregular and the h_t^* 's are presented in figure 8 (i). Estimating the heteroscedastic model, that is with measurement error variances proportional to the h_t^* 's gives an AIC of 1163 (treating the h_t^* 's as given). The resulting spline, shown in figure 8 (ii), is not too different to the one shown in 6 but the confidence interval is much narrower at the beginning and end. The smoothed irregular component in 8 (iii) is now closer to being homoscedastic. As can be seen in the fourth panel, the allowance for heteroscedasticity makes the weighting pattern at time $\tau_{105} = 35.6$ even more asymmetric.

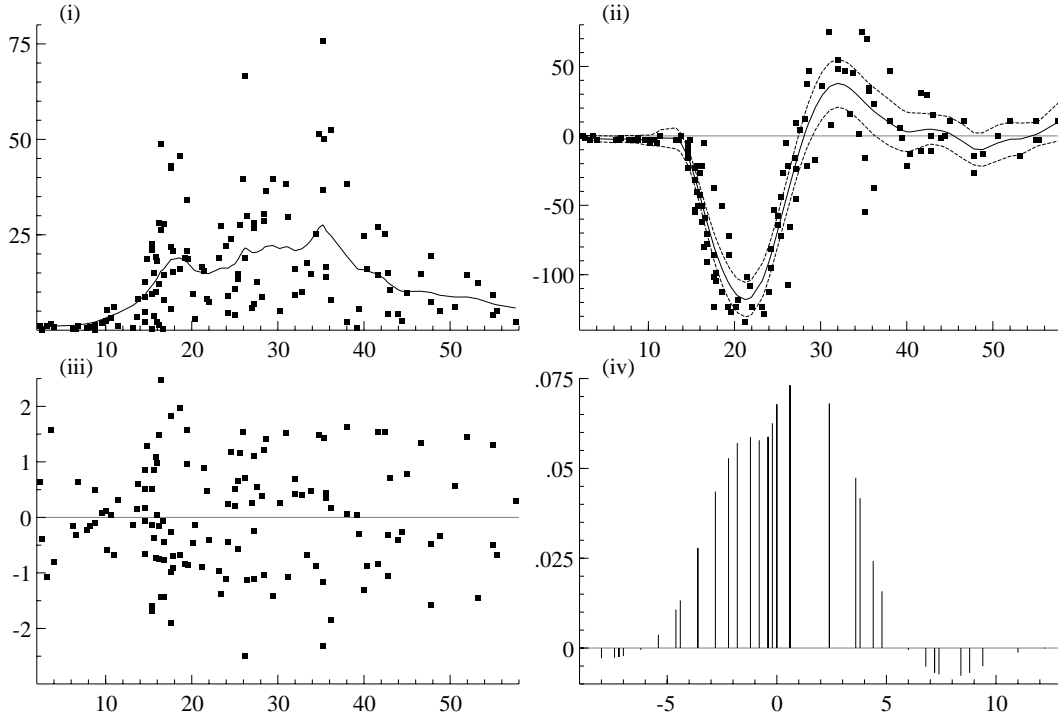


Figure 8: Motorcycle acceleration data analysed by a cubic spline corrected for heteroscedasticity. (i) absolute values of smoothed irregular and h_t^* , (ii) data with signal and confidence intervals, (iii) standardised irregular and (iv) the weights for the signal at time $\tau_{105} = 35.6$ against time distance $\tau_{105} - \tau_j$.

9 Conclusion: Back to Basics

A time series model provides a way of weighting current and past observations so as to provide good *predictions*. If the model is expressed in unobserved components it also provides a *description* of the data based on a suitable weighting of observations around a particular point. Unobserved components models with mutually uncorrelated disturbances produce weighting patterns for prediction and signal extraction which are based on sound statistical principles and accord with common sense. Three main

points emerge. The first is that signal extraction filters are symmetric in the middle of the series. The second is that it is unnecessary to consider parameter restrictions other than the obvious ones of requiring variances to be non-negative. The third is that little of practical value is lost by restricting attention to such models. (For example, the only way in which θ in the reduced form ARIMA(0, 1, 1) model for a random walk plus noise can cover the full invertibility region is by having $\rho = 1$. But given that positive values of θ yield alternate positive and negative weights in the EWMA filter, this should not be construed as an advantage.) Cases where there is a theoretical rationale for correlated disturbances, such as observations consisting of integrals of a continuous time process, are best dealt with by reference to the original underlying model.

Any ARIMA(p, d, q) model with $d = 1$ can be decomposed into a random walk and a stationary component by the BN decomposition. The filter for the random walk is one-sided but a corresponding two-sided smoother can be constructed as well. However, the generality of this procedure is somewhat illusory, since unless the ARIMA model can be obtained as the reduced form of a UC model with uncorrelated components, or can be seen to be an approximation to such a model, it is not clear that the decomposition is a particularly useful one.

Unobserved components models with heavy-tailed disturbances are robust to outliers and structural breaks. These models are nonlinear but their implied weighting patterns can be obtained approximately by matching them with Gaussian model with heteroscedastic disturbances. The weighting patterns display local asymmetries which depend on the position and magnitude of outlying observations and/or structural breaks.

The weighting patterns implied by unobserved components models may be compared with the kernels used in nonparametric time series trend estimation. The bandwidth used in nonparametric estimation plays a similar role to a signal-noise ratio in a model, and so a nonparametric approach does not necessarily imply a reduction in the number of ‘parameters’ which have to be determined. Indeed, nonparametric methods are best thought of as methods of modelling polynomials locally, but, of course, this is exactly what stochastic trends are set up to do. Perhaps the only sense in which the nonparametric methods are nonparametric is that the shape of the kernel is selected without any reference to the data, but it is difficult to see how this could possibly give any advantage with respect to robustness. The following points may be made with regard to the advantages of using UC models:

1. they can be selected and checked using standard time series methods;
2. the parameters can be estimated by ML;
3. appropriate weights are implicitly provided for points near the beginning and end of the series as well as in the middle;
4. predictions can be made together with their RMSEs;
5. they can be generalised to include other components and to deal with heteroscedasticity;
6. they can be extended to handle Poisson and other non-Gaussian distributions as in Durbin and Koopman (1999); compare with Green and Silverman (1994, chapter 5);
7. they can be made robust to outliers and structural breaks by specifying t -distributions for the disturbances; compare the robust nonparametric approach in Härdle (1990, chapter 6); and
8. by formulating a model in continuous time, the optimal weighting for irregularly spaced observations is automatically carried out.

The ability to handle irregularly spaced data means that the models are not restricted to time series. An interesting point to emerge from the example is that, in contrast to the weights used in the nonparametric approach, the model-based weights for irregularly-spaced observations are not necessarily symmetric.

Cubic splines may be seen as the smoothed estimates produced by a special case of the local linear trend model. Hence the equivalent kernel may be obtained from this model. It is surprising that the cubic spline fitting methodology, as expounded in Green and Silverman (1994) and elsewhere, makes little or no reference to the time series literature. Time series models have all the attractions listed in the previous paragraphs, while from the technical point of view, Brown and de Jong (1998) argue that the KFS is superior to the Reinsch algorithm advocated in Green and Silverman (1994).

Acknowledgments

We would like to thank Chris Chatfield, Hashem Pesaran, and participants in seminars at LSE and Cambridge for helpful comments and suggestions. The second author is a Research Fellow of the Royal Netherlands Academy of Arts and Sciences and its financial support is gratefully acknowledged.

Appendix A

The WK formula is for a doubly infinite sample and a time-invariant model. In this situation, the weights are independent of time and if it is difficult to work them out analytically they can be computed by running a smoother on a series consisting entirely of zeroes except for a one in the t -th position. Then

$$\tilde{\mu}_{t+j} = \sum_i w_i y_{t+j+i} = w_{-j}, \quad j = 0, \pm 1, \pm 2, \dots$$

so the weights appear in reverse order.

For general models in state space form, the weights can be found by the algorithm set out below. This algorithm is very general, being valid for finite samples and for models which may not be time-invariant and which may be nonstationary, noninvertible or may contain deterministic components. The calculations reported were carried out using the **SsfPack** 2.2 package of Koopman *et al* (1998) implemented for the **0x** programming language of Doornik (1998); see appendix B for a listing of the **0x** code used in section 8.1.

State space form

The future form of the Gaussian linear state space model is given by

$$\begin{aligned} y_t &= Z_t \alpha_t + G_t \varepsilon_t, & \varepsilon_t &\sim N(0, I), & t &= 1, \dots, T, \\ \alpha_{t+1} &= T_t \alpha_t + H_t \varepsilon_t, & \alpha_1 &\sim N(a, P), \end{aligned} \quad (45)$$

where y_t is the $N \times 1$ vector of observations, α_t is the $p \times 1$ unobserved state vector and ε_t is the $q \times 1$ vector of disturbances. The equation for y_t is called the measurement equation and the equation for α_{t+1} is referred to as the transition equation. The disturbances in the measurement and transition equations are uncorrelated if $H_t G_t' = 0$. The system matrices Z_t , G_t , T_t and H_t , with appropriate dimensions, are fixed. The state space model (45) is said to be time-invariant when the system matrices are constant over time, that is $Z_t = Z$, $T_t = T$, $G_t = G$ and $H_t = H$, for $t = 1, \dots, T$.

Kalman filter and smoother

The Kalman filter (KF) evaluates the MMSLE of the state vector α_{t+1} conditional on the observations $Y_t = \{y_1, \dots, y_t\}$, that is $a_{t+1|t} = E(\alpha_{t+1}|Y_t)$. The KF is given by

$$\begin{aligned} v_t &= y_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + G_t G_t', & M_t &= T_t P_{t|t-1} Z_t' + H_t G_t', \\ a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} T_t' + H_t H_t' - K_t M_t', & t &= 1, \dots, T, \end{aligned} \quad (46)$$

with Kalman gain matrix $K_t = M_t F_t^{-1}$ and with the initialisations $a_{1|0} = a$ and $P_{1|0} = P$. In the case of a time-invariant model, the KF may converge to a *steady-state* such that $P_{t|t-1} = \bar{P}$, $F_t = \bar{F}$ and $K_t = \bar{K}$, for $t = s + 1, \dots$, and some large s . The filtered estimator $a_{t|t} = E(\alpha_t|Y_t)$ can also be computed using the KF. Smoothed estimators of the state and disturbance vector, that is $a_{t|T} = E(\alpha_t|Y_T)$ and $e_{t|T} = E(\varepsilon_t|Y_T)$, are evaluated using the backwards smoothing filter

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \quad t = T, \dots, 1, \quad (47)$$

where $L_t = T_t - K_t Z_t$ and with initialisations $r_T = 0$ and $N_T = 0$. For example, $a_{t|T} = a_{t|t-1} + P_{t|t-1} r_{t-1}$ and $e_{t|T} = G_t' F_t^{-1} v_t + J_t' r_t$ where $J_t = H_t - K_t G_t$. The recursion for N_{t-1} is required for computing smoothed variances. More details can be found in Koopman (1998).

Computing Weights

We now show how to compute the weights assigned to observations when carrying out prediction, filtering and smoothing. We have $a_{t|t-1} = \sum_{j=1}^{t-1} w_j(a_{t|t-1}) y_j$, $a_{t|t} = \sum_{j=1}^t w_j(a_{t|t}) y_j$, $a_{t|T} = \sum_{j=1}^T w_j(a_{t|T}) y_j$ and $e_{t|T} = \sum_{j=1}^T w_j(e_{t|T}) y_j$. Further, we have $s_{t|t-1} = Z_t a_{t|t-1}$, $s_{t|t} = Z_t a_{t|t}$ and $s_{t|T} = Z_t a_{t|T}$. For prediction and filtering we apply the KF up to time t and then implement the backwards recursion

$$w_j(\cdot) = B_{t,j} K_j, \quad B_{t,j-1} = B_{t,j} T_j - w_j(\cdot) Z_j, \quad j = t-1, t-2, \dots, 1, \quad (48)$$

with the initialisation $B_{t,t-1}$ reported in table 3. For smoothing we apply the KF up to time T and the backwards smoothing recursion (47) up to time t . Consequently, the backwards recursion (48) is used for computing the “past” weights and for the “future” weights we apply the forwards recursion

$$w_j(\cdot) = B_{t,j}^* C_j, \quad B_{t,j+1}^* = B_{t,j}^* L_j', \quad j = t+1, \dots, T, \quad (49)$$

where $C_j = Z_j' F_j^{-1} - L_j' N_j K_j$ and with the initialisation $B_{t,t+1}^*$ reported in table 3. Proofs, further details and some illustrations are given by Koopman and Harvey (1999).

Table 3: Algorithm for computing weights

\cdot	operation	initialisation $B_{t,t-1}$ in (48)	weight $w_t(\cdot)$	initialisation $B_{t,t+1}^*$ in (49)
$a_{t t-1}$	filtering (pred)	I	0	0
$a_{t t}$	filtering (cont)	$I - P_{t t-1} U_t$	$P_{t t-1} Z_t' F_t^{-1}$	0
$a_{t T}$	smoothing	$I - P_{t t-1} N_{t-1}$	$P_{t t-1} C_t$	$P_{t t-1} L_t'$
$s_{t t-1}$	filtering (pred)	Z_t	0	0
$s_{t t}$	filtering (cont)	$G_t G_t' F_t^{-1} Z_t$	$I - G_t G_t' F_t^{-1}$	0
$s_{t T}$	smoothing	$G_t G_t' C_t'$	$Z_t P_{t t-1} C_t$	$Z_t P_{t t-1} L_t'$
$e_{t T}$	smoothing	$-(G_t' C_t' + H_t' N_t L_t)$	$G_t' F_t^{-1} - J_t' N_t K_t$	J_t'

Appendix B

The listing of the `Ox` code for the computations of section 8.1 is given below. The motorcycle acceleration data and the `Ox` programs used for all computations in this paper can be found on the Internet:

<http://center.kub.nl/staff/koopman>

The function `MaxLik()` estimates the smoothing parameter γ and the function `DrawComponents()` generates the output used to draw figure 6.

```
#include <oxstd.h>
#import <maximize>
#include <packages/ssfpack/ssfpack.h>

static decl s_mY, s_mX, s_mD, s_cT;
static decl s_q, s_dSigma;

SetSplParameters(const vP)
{
    s_q = exp(2. * vP[0]);
}

SplLogLikc(const vY, const pdLik, const pdVar)
{
    decl mphi, momega, msigma, mj_phi=<>, mj_omega=<>, mx=<>, ret_val;

    GetSsfSpline(s_q, s_mD, &mphi, &momega, &msigma, &mj_phi, &mj_omega, &mx);
    ret_val = SsfLikConc(pdLik, pdVar, vY, mphi, momega, msigma,
        <>, mj_phi, mj_omega, <>, mx);
    s_dSigma = sqrt(pdVar[0]); // get sigma from SsfLikConc
    return ret_val; // 1 indicates success, 0 failure
}

Likelihood(const vP, const pdLik, const pvSco, const pmHes)
{
    // arguments dictated by MaxBFGS()
    decl dvar, ret_val;

    SetSplParameters(vP); // map vP to AR(1) model
    ret_val = SplLogLikc(s_mY, pdLik, &dvar);
    pdLik[0] /= s_cT; // log-likelihood scaled by sample size
    return ret_val; // 1 indicates success, 0 failure
}

SplStderr(const vP)
{
    decl covar, invcov, dsig = s_dSigma, dq = s_q, result;

    result = Num2Derivative(Likelihood, vP, &covar);
    s_dSigma = dsig, s_q = dq; // reset after Num2Der

    if (!result)
    { print("Covar() failed in numerical second derivatives\n");
      return zeros(vP);
    }
    invcov = invertgen(-covar, 30);
    return sqrt(diagonal(invcov) / s_cT)';
}
```

```

MaxLik()
{
    decl vp, ir, dlik;
    vp = <-2>; // set starting values
    MaxControl(-1, 5, 1);
    MaxControlEps(1e-7, 1e-5); // tighter convergence criteria
    ir = MaxBFGS(Likelihood, &vp, &dlik, 0, TRUE);

    println("\n", MaxConvergenceMsg(ir),
        " using numerical derivatives",
        "\nLog-likelihood = ", "%.8g", dlik * s_cT,
        "; variance = ", s_dSigma, " (= dVar); n=", s_cT);
    print("parameters [transformed] with (standard errors):",
        "%cf", {"%12.5g", "%12.5g", " (%7.5f)"}, exp(vp[0]) ~ vp ~ SplStderr(vp));
}

DrawComponents(const loc)
{
    decl cst, mphi, momega, msigma, mj_phi=<>, mj_omega=<>, mx=<>, ret_val, md;

    GetSsfSpline(s_q, s_mD, &mphi, &momega, &msigma, &mj_phi, &mj_omega, &mx);
    cst = columns(mphi);

    SsfMomentEst(ST_SMO, &md, s_mY, mphi, momega, msigma,
        <>, mj_phi, mj_omega, <>, mx);
    // smoothed state vector: plot some vectors from md

    SsfMomentEst(DS_SMO, &md, s_mY, mphi, momega, msigma,
        <>, mj_phi, mj_omega, <>, mx);
    // smoothed disturbance vector: plot some vectors from md

    decl index = range(-20, 20);
    decl mw = SsfWeights(ST_SMO, loc, s_mY, mphi, momega, msigma,
        <>, mj_phi, mj_omega, <>, mx);
    // plot weights against time (s_mX[0][loc + index - 1]) and against index
}

main()
{
    decl myt = loadmat("acc.in7"); // load motorcycle data

    s_mX = myt[0] []; // time in milliseconds (tau_t)
    s_mY = myt[1] []; // acceleration (y_t)
    s_mD = myt[2] []; // delta_t = 5 (tau_t+1 - tau_t)
    s_cT = columns(s_mY); // no of observations

    MaxLik(); // maximum likelihood estimation
    DrawComponents(105); // signal and weights for t=105
}

```

References

- Bell, W.R. (1984) Signal extraction for nonstationary time series, *Annals of Statistics*, 13, pp.646-664.
- Beveridge, S, and C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'Business Cycle'. *Journal of Monetary Economics*, 7, pp.151-174.
- Box, G.E.P. and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, revised ed. San Francisco: Holden-Day.
- Brown, P.E. and P. de Jong (1998) The Reinsch algorithm and the Kalman filter smoother. Mimeo, Statistics Department, LSE.
- Burrige, P. and K.F. Wallis (1988) Prediction theory for autoregressive-moving average processes. *Econometric Reviews*, 7, pp.65-69.
- Doornik, J.A. (1998) *Object-Oriented Matrix Programming using Ox 2.0*. London: Timberlake Consultants Press.
- Durbin, J. and S.J. Koopman (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika*, 84, pp.669-84.
- Durbin, J. and S.J. Koopman (1999) Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society, Series B*, forthcoming.
- Ghysels, E., A.C. Harvey and E. Renault (1996) Stochastic Volatility. In C.R. Rao and G.S. Maddala (eds.) *Statistical Methods in Finance*. Amsterdam: North-Holland, pp.119-191.
- Gijbels, I., A. Pope and M.P. Wand (1999) Understanding exponential smoothing via kernel regression. *Journal of the Royal Statistical Society, Series B*, 61, pp.39-50.
- Green, P.G. and B.W. Silverman (1994) *Nonparametric regression and generalized linear models*. London: Chapman and Hall.
- Härdle, W. (1990) *Applied nonparametric regression*. Cambridge: Cambridge University Press.
- Hart, J.D. (1994) Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society, Series B*, 56, pp.529-42.
- Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A.C. and A. Jaeger (1993) Detrending, stylised facts and the business cycle. *Journal of Applied Econometrics*, 8, pp.231-47.
- Hillmer, S.C. and G.C. Tiao (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77, pp.63-70.
- Hodrick, R.J. and E.C. Prescott (1997) Postwar US business cycles: an empirical investigation, *Journal*

of *Money, Credit and Banking*, 24, pp.1-16.

Kohn, R., C.F. Ansley and C-H. Wong (1992) Nonparametric spline regression with autoregressive moving average errors. *Biometrika*, 79, pp.335-46.

Koopman, S.J. (1998) Kalman filtering and smoothing. In P. Armitage and T. Colton (eds), *The Encyclopedia of Biostatistics*, Chichester: Wiley and Sons.

Koopman, S.J. and A.C. Harvey (1999) Computing Observation Weights for Signal Extraction and Filtering. CentER Discussion paper, Tilburg University.

Koopman, S.J., N. Shephard and J. Doornik (1998) Statistical algorithms for models in state space using SsfPack 2.2. CentER Discussion paper, Tilburg University.

Ord, J.K, A.B. Koehler and R.D. Snyder (1997) Estimation and prediction for a class of dynamic nonlinear statistical model, *Journal of the American Statistical Association*, 92, pp.1621-1629.

Proietti, T. and A.C.Harvey (1999), A Beveridge-Nelson Smoother, mimeo.

Riani, M (1998) Weights and Robustness of Model-based Seasonal Adjustment. *Journal of Forecasting*, 17, pp.19-34.

Sandmann, G. and S.J. Koopman (1998) Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood. *Journal of Econometrics*, 87, pp.271-301.

Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika*, 81, pp.115-131.

Shephard, N. and M.K. Pitt (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84, pp.653-667.

Silverman, B.W. (1984) Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, 12, pp.898-916.

Silverman, B.W. (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society, Series B*, 47, pp.1-52.

Snyder, R.D. (1985) Recursive estimation of dynamic linear statistical models. *Journal of the Royal Statistical Society, Series B*, 47, pp.272-276.

Watson, M. (1986) Univariate detrending methods with stochastic trends. *Journal of Monetary Economics*, 18, pp.49-75.

Wecker, W.E. and C.F. Ansley (1983) The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78, pp.81-89.

Whittle, P. (1983) *Prediction and Regulation*, 2nd ed. Oxford: Blackwell.